Overview of the CLEF-2023 CheckThat! Lab Task 5 on Authority Finding in Twitter

Fatima Haouari[1], Tamer Elsayed[1] and Zien Sheikh Ali[1]

[1]*Qatar University, Doha, Qatar*

Abstract

We present an overview of Task 5 of the sixth edition of the CheckThat! Lab, which is a part of the 2023 Conference and Labs of the Evaluation Forum (CLEF). In this Authority Finding task, participating systems are required to retrieve a set of authority Twitter accounts for a given rumor expressed in a propagating Arabic tweet. Two teams participated in this first version of the task, submitting a total of four runs, 3 of which managed to achieve significant improvement over the baseline. In this paper, we present our data collection approach, evaluation setup, and an overview of the participating systems. We release to the research community the dataset as well as the evaluation scripts, which should enable further research on this new task.

Keywords

Expert Finding, Rumor Verification, Fact-Checking, Veracity, Social Media

1. Introduction

The explosive growth of social media platforms over the recent years has facilitated the spread of misinformation and fake news. To address this issue, a myriad of initiatives addressed verification of claims propagating on social media [1, 2, 3], or over the Web [4, 5]. Additionally, a plethora of fact-checking organizations were launched worldwide covering multiple regions and languages to perform manual claim verification. However, given the large scale of misinformation circulating in different communication platforms, there is still an urgent need to debunk it despite all the efforts made.

To extend existing efforts and foster system development to debunk misinformation, CheckThat! Lab was launched in 2018 featuring a variety of tasks to help automate the fact-checking process. The CheckThat! lab runs for the sixth time under the umbrella of CLEF 2023.[1] The objective of this edition of the lab [6] was to encourage work on detecting multimodal check-worthy claims, detecting subjectivity in news articles, predicting the political bias of news articles and news media, predicting the factuality of reporting of news media, and finding authorities that can help verify rumors in Twitter.

In this paper, we describe in detail Task 5 of this year's lab,[2] *Authority Finding in Twitter*. Task 5 is defined as follows: "*Given a tweet stating a rumor, a model has to retrieve a ranked list*

CLEF 2023: Conference and Labs of the Evaluation Forum, September 18-21, 2023, Thessaloniki, Greece

200159617@qu.edu.qa (F. Haouari); telsayed@qu.edu.qa (T. Elsayed); zs1407404@qu.edu.qa (Z. S. Ali)

[1] https://checkthat.gitlab.io/

[2] Refer to [7] for an overview of the full CheckThat! 2023 lab.

of authority Twitter accounts that can help in verifying the rumor; i.e. they may tweet evidence that supports or denies that rumor [8]". An illustration of the task with an example from our data[3] is presented in Figure. 1.

The rest of this paper is organized as follows. We give an overview of the related work in Section 2. Section 3 introduces our task dataset. Section 4 discusses the reported evaluation measures. We then present an overview of the participants' systems in Section 5, and discuss the evaluation results in Section 6. A failure analysis discussion is presented in Section 7. Finally, we conclude in Section 8.

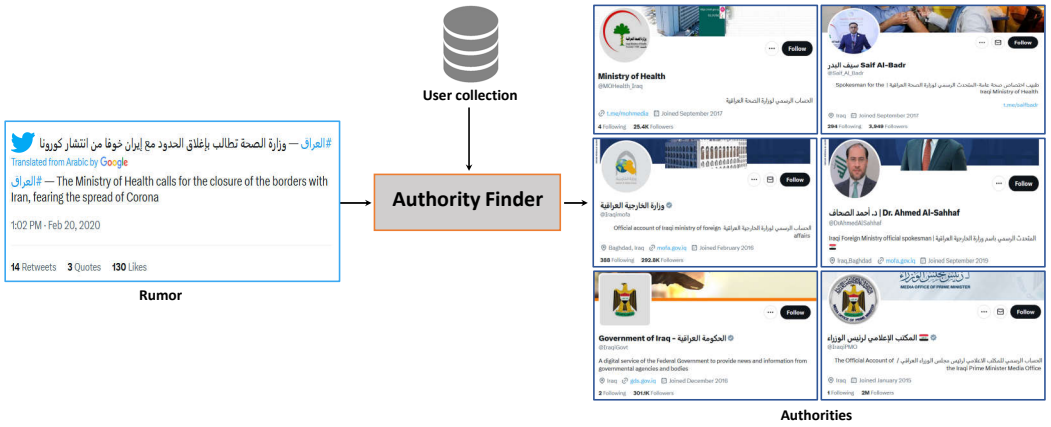

Figure 1: An illustration of the *Authority Finding in Twitter* task with a real-world example.

2. Related Work

This section reviews the literature related to debunking rumors in social media (Section 2.1), and expert and authority finding in social media research studies (Section 2.2).

2.1. Debunking Rumors in Social Media

Several studies suggested exploiting online debunkers. i.e, users who have a tendency to correct rumors by sharing fact-checking URLs [9, 10, 11, 12]. Vo and Lee [9, 10] and You et al. [11] introduced fact-checking URLs recommender systems for online debunkers in Twitter to support them in correcting misinformation. Vo and Lee [13] leveraged the conversation threads of online debunkers in Twitter, and proposed a fact-checking response generator system. Vo and Lee [12] also proposed a framework to retrieve relevant fact-checking articles and integrate them into rumor spreaders conversation threads to discourage rumors spreading in social media. In this work, we consider authorities as credible debunkers who can either support or debunk a rumor circulating in Twitter [8].

[3]That rumor has 14 authorities as per the annotators; only six are shown in the figure.

2.2. Expert and Authority Finding in Social Media

There exists a considerable body of literature on expert finding targeting many domains [14], focusing on community question answering platforms [15, 16], academic social networking sites [17], or social media platforms [18, 19, 20, 21, 22, 23]

The literature review shows that expert finding studies in social media can be grouped into three main categories. The first category covers studies targeting topic expert finding, where given a topic the task is to retrieve a ranked list of expert accounts [18, 19]. Second, studies focusing on local expert finding, which is similar to topic expert finding, but the difference is that the experts should be within a specific location [20, 21]. The third category which is the most relevant to our task is misinformation-based expert finding, where the task is given a rumor in social media, retrieve a ranked list of experts accounts [22, 23].

The majority of prior research on expert finding leveraged Twitter lists [20, 18, 19, 23] to represent users as suggested by Ghosh et al. [18] who showed that it plays a significant role to find topic experts. Additionally, several methods were proposed to address the task adopting either the Bayes theorem for estimating the probability that a user is an expert [22, 23], proposing a query-dependent personalized PageRank [19], or utilizing a learning-to-rank framework [20, 21].

Differently, Haouari et al. [24] introduced the task of authority finding for rumor verification in Twitter where not all experts are authorities. The authors considered the task as a sub-problem of topic expert finding in social media. They released the first authority finder test collection targeting the Arabic language, and proposed a hybrid model that incorporates the lexical and semantic textual representation of users, and users networks features.

3. Authority Finding Dataset

Training Data We adopted AuFIN [24] test collection which comprises 150 rumors expressed in tweets, associated with 1,044 authority Twitter accounts, and a user collection of 395,231 Twitter accounts along with their Twitter lists (1,192,284 unique lists). Each authority is graded whether she is *highly relevant* or *relevant* to the rumor, i.e., having a higher priority to be contacted for verification or not. The rumors cover three categories: politics, sports, and health (50 from each category). We split the rumors into 120 for training and 30 for development.

Testing Data We collected 30 new rumors from AraFacts [25], where we focused on the ones collected from Misbar and Fatabayyano which were exploited recently to construct Arabic rumor verification [26] and fake news detection [27] datasets. We selected 10 rumors from each one of the three above categories. For each rumor, two annotators were asked separately to identify all possible authority Twitter accounts who can help in supporting or debunking that rumor following the annotation guidelines and data quality evaluation provided by Haouari et al. [24]. Cohen's Kappa inter-annotator agreement [28] was 0.91 and 0.42 for the authority label and the graded relevance respectively, which correspond to almost perfect and moderate agreements respectively [29].

Table 1 presents an overall statistics of the rumor collection and the relevance judgments for each data split. For an overall summary of the user collection, we refer the reader to Haouari

et al. [24]. An example for a test rumor along with its corresponding authorities and their associated relevance grades is presented in Table 2.

Table 1
Rumor collection and relevance judgments statistics.

Data split	Rumors	Authorities
Training	120	849
Development	30	195
Testing	30	172

Table 2
An example of a test rumor with its corresponding authorities.

Rumor Tweet: A while ago, the United Nations vehicles evacuated the #Tunisian embassy and transferred the #ambassador and the #embassy employees to a safe place.#Libya		
	Authority	**Relevance**
1.	**(@UNSMILibya)** The Official Twitter Account of UNSMIL. The United Nations Support Mission in Libya	highly relevant
2.	**(@UNNewsArabic)** The Official Twitter Account of the United Nations News	highly relevant
3.	**(@TunisieDiplo)** Tunisian Ministry of Foreign Affairs and Immigration	highly relevant
4.	**(@OJerandi)** Tunisian Minister of Foreign Affairs and Immigration	highly relevant
5.	**(@TnPresidency)** Official Account of the Tunisian Presidency	relevant

4. Evaluation Measures

We report P@1 and P@5 to evaluate how well the participating systems retrieve Twitter authorities at the top of a short retrieved ranked list. We further report NDCG@5 to measure the ability of systems to retrieve highly relevant authority Twitter accounts higher up in that list. We adopt P@5 as the official evaluation measure.

5. Overview of the Participating Systems

A total of two teams participated in this task, submitting four runs. We present below an overview of each team approach.

bigIR bigIR team adopted the hybrid model proposed by Haouari et al. [24] that incorporates lexical, semantic, and user network features to retrieve authorities. To get the initial candidates, they first utilized the bio_lists_index[4] to get the lexical scores of the top 1000 users, and then used the network features, i.e., count of Twitter lists, followers, and followees, to compute the initial scores. To get the semantic scores for the top 100 initial candidates, they used Arabic BERT [30] fine-tuned with the full training data and deployed in Tahaqqaq real-time system [31]. Finally, the initial and the semantic scores were interpolated to get the hybrid final scores for

the candidates to rerank them. The team submitted three runs varying the interpolation weight α to 0.5, 0.75, and 0.25 for Hybrid1, Hybrid2, and Hybrid3 runs respectively. In other words, they assigned equal weights to both scores, more weight to the initial scores, and more weight to the semantic scores for the runs respectively.

ES-VRAI [32] Similar to bigIR, they leveraged the lexical matching and user features to retrieve authorities, however they neglected the semantic matching between the rumor and the users. The team utilized the bio_lists_index [24] and optimised BM25 hyper-parameters to retrieve the initial candidates. They then re-scored and reranked the initially retrieved candidates by exploiting the users' metadata features.

6. Evaluation Results

Table. 3 shows the official results for Task 5. We can clearly notice that only three runs, which are submitted by bigIR team, managed to outperform the BM25 baseline. We observe that bigIR (Hybrid3) run is the top performing in terms of all evaluation measures. Contrary to the findings presented by Haouari et al. [24], which states "the initial retrieval score is more indicative than the semantic score," bigIR (Hybrid3) assigns higher weight (0.75 vs. 0.25) to the semantic scores than to the initial retrieval scores. Moreover, bigIR (Hybrid2), which assigns higher weight (0.75) to the initial scores is the least performing compared to other bigIR runs. These results highlight that there is no best setup for setting the interpolation weight, and that the data plays a significant factor, i.e., α could vary for different data to achieve the best performance. Additionally, we can conclude that the ArabicBERT model fine-tuned on the training data is able to generalize to new emerging data.

Table 3
Official evaluation results, in terms of P@1, P@5, and nDCG@5. The teams are ranked by the official evaluation measure P@5. Here, *BM25 baseline* is a lexical retrieval model where the query is the rumor text and the index is bio+lists released by Haouari et al. [24]. Submissions with a + sign indicate submissions by task organisers.

	Team (run ID)	**Evaluation Scores**		
		P@5	**P@1**	**nDCG@5**
1	+bigIR (Hybrid3)	0.260	0.367	0.297
2	+bigIR (Hybrid1)	0.247	0.367	0.282
3	+bigIR (Hybrid2)	0.227	0.333	0.247
	BM25 baseline	0.087	0.133	0.104
4	ES-VRAI (Model1)	0.067	0.067	0.071

[4] https://github.com/Fatima-Haouari/AuFIN

7. Failure Analysis

We conducted a thorough failure analysis on all rumors on which the best model, bigIR (Hybrid3), could not retrieve any authority at depth 100. This comprises 5 rumors of our test set (16.7%). We study both the *false negatives*, and *false positives* error types. This section presents the main potential reasons of these failures in detail.

Table 4
False negatives cases of (translated) rumors and user representations. The misspelled entities, and non-Arabic keywords are in red, and blue respectively. Underlined are the entities mentioned in the rumor.

Tweet (Failure Reason(s))	**Non-retrieved Authority & Textual Representation**
T_1 **(Lack of context, No lexical overlap, Misspelling, non-Arabic keywords)** Urgent the dismissal of Djamel Belmadi and the appointment of Samir Zaoui #Algerie #Algeriz #Algeria	**@FAFAlgeria** Algerian Football Federation Algerian Football Clubs Algeria List 2021 Afcon Feet Official Official Accounts of African Football Federations Algerian National Team Football Football Sport News Football Sport News Algerian National Team...
T_2 **(Lack of context, No lexical overlap)** In a study, it says that you should not exercise after taking the vaccine until at least 3 weeks later. #vaccine_is_compulsory	**@WHOEMRO** The official Twitter account of the WHO Regional Office for the Eastern Mediterranean Who is WHO Regional accounts and accounts of the World Customs Organization...accounts containing information about the epidemics COVID-19 Coronavirus updates...

7.1. False Negatives

Table 4 presents examples of rumors along with authorities failed to be retrieved at depth 100. We discuss below the potential reasons behind these failures.

1. **No lexical overlap:** Some rumors do not have any lexical overlap with the user textual representation, e.g., T_1 and T_2. Thus, the model fails to retrieve these users, as the model initial candidates are retrieved by lexical matching.

2. **Lack of context:** Some rumors may require context expansion to retrieve relevant authorities, e.g., T_1, which is about the appointment of a new coach to train the Algerian football team, where it is not clear that Djamel Belmadi is a football coach. Similarly, for T_2, some additional context is required to understand that the vaccine is for COVID-19.

3. **Misspelled entities:** Some rumors have misspelled entities, which may affect the retrieval of relevant authorities, e.g., in T_1, Algeria is misspelled (the only one in Arabic in this tweet, indicated in red).

4. **Non-Arabic keywords:** Some tweets may mention entities in non-Arabic, e.g., in T_1. The model used by bigIR preprocesses the rumors by excluding non-Arabic keywords, while these can be key terms in the rumor.

Table 5
False positives cases of (translated) rumors and user representations. The lexical overlap is in green. Underlined are the entities mentioned in the rumor.

Tweet (Failure Reason(s))	Retrieved Authority & Textual Representation
T_3 **(Lack of context)** Urgent the dismissal of Djamel Belmadi and the appointment of Samir Zaoui #Algerie #Algeriz #Algeria	**@OmaniDecisions** Appointments an account concerned with news of appointments in ministries, authorities, committees and government companies Government accounts Oman Official government accounts...
T_4 **(Lack of context)** In a study, it says that you should not exercise after taking the vaccination until at least 3 weeks later. #vaccination_is_compulsory	**@drshamsah** Consultant pediatrician and pediatric emergency Mubarak Al-Kabeer, President of the College of Pediatrics... vaccination, compulsory vaccination...

7.2. False Positives

In Table 5, we present examples of rumors along with their top retrieved non-authority. We found that the main reason behind this type of failure is lack of context, which leads to retrieving *false positive* users just because of some lexical overlap on some of the rumor keywords. For example, T_3 and T_4, as discussed in Section 7.1, need additional context. This led the model to retrieve irrelevant users who have lexical overlap on some of the rumor keywords.

8. Conclusion

In this paper, we presented a detailed overview of the CLEF 2023 CheckThat! Lab Task 5 for authority finding in Twitter. Some of the participant systems leveraged the lexical matching along with user network features only, while others expanded this by incorporating the semantic matching between the rumor and the user textual representations, which showed a significant improvement over lexical matching. Moreover, all systems adopted the user profile and their Twitter lists to represent users, which is the SOTA user representation for this task.

Acknowledgments

The work of Fatima Haouari was supported by GSRA grant #GSRA6-1-0611-19074 from the Qatar National Research Fund (a member of Qatar Foundation). The work of Tamer Elsayed was made possible by NPRP grant #NPRP-11S-1204-170060 from the Qatar National Research Fund. The statements made herein are solely the responsibility of the authors.

References

[1] G. Gorrell, E. Kochkina, M. Liakata, A. Aker, A. Zubiaga, K. Bontcheva, L. Derczynski, SemEval-2019 task 7: RumourEval, Determining Rumour Veracity and Support for Rumours, in: Proceedings of the 13th International Workshop on Semantic Evaluation, Association for Computational Linguistics, Minneapolis, Minnesota, USA, 2019, pp. 845–854.

[2] P. Nakov, D. S. M. Giovanni, T. Elsayed, A. Barrón-Cedeño, R. Míguez, S. Shaar, F. Alam, F. Haouari, M. Hasanain, W. Mansour, B. Hamdan, Z. S. Ali, N. Babulkov, A. Nikolov, G. K. Shahi, J. M. Struß, T. Mandl, M. Kutlu, Y. S. Kartal, Overview of the CLEF-2021 CheckThat! lab on detecting check-worthy claims, previously fact-checked claims, and fake news, CLEF 2021, 2021.

[3] A. Barrón-Cedeño, T. Elsayed, P. Nakov, G. Da San Martino, M. Hasanain, R. Suwaileh, F. Haouari, N. Babulkov, B. Hamdan, A. Nikolov, et al., Overview of CheckThat! 2020: Automatic Identification and Verification of Claims in Social Media, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction: 11th International Conference of the CLEF Association, CLEF 2020, Thessaloniki, Greece, September 22–25, 2020, Proceedings 11, Springer, 2020, pp. 215–236.

[4] F. Haouari, Z. S. Ali, T. Elsayed, bigIR at CLEF 2019: Automatic Verification of Arabic Claims over the Web., 2019.

[5] T. Elsayed, P. Nakov, A. Barrón-Cedeno, M. Hasanain, R. Suwaileh, G. Da San Martino, P. Atanasova, Overview of the CLEF-2019 CheckThat! Lab: Automatic Identification and Verification of Claims, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction: 10th International Conference of the CLEF Association, CLEF 2019, Lugano, Switzerland, September 9–12, 2019, Proceedings 10, Springer, 2019, pp. 301–321.

[6] A. Barrón-Cedeño, F. Alam, T. Caselli, G. Da San Martino, T. Elsayed, A. Galassi, F. Haouari, F. Ruggeri, J. M. Struss, R. N. Nandi, G. S. Cheema, D. Azizov, P. Nakov, The CLEF-2023 CheckThat! Lab: Checkworthiness, Subjectivity, Political Bias, Factuality, and Authority, in: J. Kamps, L. Goeuriot, F. Crestani, M. Maistro, H. Joho, B. Davis, C. Gurrin, U. Kruschwitz, A. Caputo (Eds.), Advances in Information Retrieval, Springer Nature Switzerland, Cham, 2023, pp. 506–517.

[7] A. Barrón-Cedeño, F. Alam, A. Galassi, G. Da San Martino, P. Nakov, T. Elsayed, D. Azizov, T. Caselli, G. Cheema, F. Haouari, M. Hasanain, M. Kutlu, C. Li, F. Ruggeri, J. M. Struß, W. Zaghouani, Overview of the CLEF–2023 CheckThat! Lab checkworthiness, subjectivity, political bias, factuality, and authority of news articles and their source, in: A. Arampatzis, E. Kanoulas, T. Tsikrika, S. Vrochidis, A. Giachanou, D. Li, M. Aliannejadi, M. Vlachos,

G. Faggioli, N. Ferro (Eds.), Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Fourteenth International Conference of the CLEF Association (CLEF 2023), 2023.
[8] F. Haouari, T. Elsayed, Detecting Stance of Authorities Towards Rumors in Arabic Tweets: A Preliminary Study, in: Advances in Information Retrieval, Springer Nature Switzerland, Cham, 2023, pp. 430–438.
[9] N. Vo, K. Lee, The Rise of Guardians: Fact-Checking URL Recommendation to Combat Fake News, in: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR '18, Association for Computing Machinery, New York, NY, USA, 2018, p. 275–284.
[10] N. Vo, K. Lee, Standing on the Shoulders of Guardians: Novel Methodologies to Combat Fake News, in: Disinformation, Misinformation, and Fake News in Social Media, Springer, 2020, pp. 183–210.
[11] D. You, N. Vo, K. Lee, Q. LIU, Attributed Multi-Relational Attention Network for Fact-Checking URL Recommendation, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 1471–1480.
[12] N. Vo, K. Lee, Where Are the Facts? Searching for Fact-Checked Information to Alleviate the Spread of Fake News, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020, pp. 7717–7731.
[13] N. Vo, K. Lee, Learning from Fact-Checkers: Analysis and Generation of Fact-Checking Language, in: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2019, pp. 335–344.
[14] N. Nikzad-Khasmakhi, M. Balafar, M. R. Feizi-Derakhshi, The State-of-the-Art in Expert Recommendation Systems, Engineering Applications of Artificial Intelligence 82 (2019) 126–147.
[15] N. Nikzad-Khasmakhi, M. Balafar, M. R. Feizi-Derakhshi, C. Motamed, BERTERS: Multimodal Representation Learning for Expert Recommendation System with Transformers and Graph Embeddings, Chaos, Solitons & Fractals 151 (2021) 111260.
[16] Z. Fallahnejad, H. Beigy, Attention-Based Skill Translation Models for Expert Finding, Expert Systems with Applications 193 (2022) 116433.
[17] D. Wu, S. Fan, F. Yuan, Research on Pathways of Expert Finding on Academic Social Networking Sites, Information Processing & Management 58 (2021) 102475.
[18] S. Ghosh, N. Sharma, F. Benevenuto, N. Ganguly, K. Gummadi, Cognos: Crowdsourcing Search for Topic Experts in Microblogs, in: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '12, Association for Computing Machinery, New York, NY, USA, 2012, p. 575–590.
[19] P. Lahoti, G. De Francisci Morales, A. Gionis, Finding Topical Experts in Twitter via Query-Dependent Personalized PageRank, in: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, 2017, pp. 155–162.
[20] W. Niu, Z. Liu, J. Caverlee, On Local Expert Discovery via Geo-Located Crowds, Queries, and Candidates, ACM Trans. Spatial Algorithms Syst. 2 (2016).
[21] W. Niu, Z. Liu, J. Caverlee, LExL: A Learning Approach for Local Expert Discovery on Twitter, in: Advances in Information Retrieval: 38th European Conference on IR Research,

ECIR 2016, Padua, Italy, March 20–23, 2016. Proceedings 38, Springer, 2016, pp. 803–809.
[22] C. Liang, Z. Liu, M. Sun, Expert Finding for Microblog Misinformation Identification, in: Proceedings of COLING 2012: Posters, 2012, pp. 703–712.
[23] G. Li, M. Dong, F. Yang, J. Zeng, J. Yuan, C. Jin, N. Q. V. Hung, P. T. Cong, B. Zheng, Misinformation-Oriented Expert Finding in Social Networks, World Wide Web 23 (2020) 693–714.
[24] F. Haouari, T. Elsayed, W. Mansour, Who can verify this? finding authorities for rumor verification in Twitter, Information Processing & Management 60 (2023) 103366.
[25] Z. S. Ali, W. Mansour, T. Elsayed, A. Al-Ali, AraFacts: The First Large Arabic Dataset of Naturally Occurring Claims, in: Proceedings of the Sixth Arabic Natural Language Processing Workshop, 2021, pp. 231–236.
[26] F. Haouari, M. Hasanain, R. Suwaileh, T. Elsayed, ArCOV19-Rumors: Arabic COVID-19 Twitter Dataset for Misinformation Detection, in: Proceedings of the Sixth Arabic Natural Language Processing Workshop, 2021, pp. 72–81.
[27] A. Khalil, M. Jarrah, M. Aldwairi, Y. Jararweh, Detecting Arabic fake news using machine learning, in: Proceedings of the International Conference on Intelligent Data Science Technologies and Applications, IDSTA '21, 2021, pp. 171–177.
[28] J. Cohen, A coefficient of agreement for nominal scales, Educational and psychological measurement 20 (1960) 37–46.
[29] J. R. Landis, G. G. Koch, The measurement of observer agreement for categorical data, Biometrics (1977) 159–174.
[30] A. Safaya, M. Abdullatif, D. Yuret, KUISAIL at SemEval-2020 task 12: BERT-CNN for offensive speech identification in social media, in: Proceedings of the Fourteenth Workshop on Semantic Evaluation, International Committee for Computational Linguistics, Barcelona (online), 2020, pp. 2054–2059.
[31] Z. Sheikh Ali, W. Mansour, F. Haouari, M. Hasanain, T. Elsayed, A. Al-Ali, Tahaqqaq: A Real-Time System for Assisting Twitter Users in Arabic Claim Verification, in: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2023.
[32] H. T. Sadouk, F. Sebbak, H. E. Zekiri, ES-VRAI at CheckThat! 2023: Leveraging Bio and Lists Information for Enhanced Rumor Verification in Twitter, in: M. Aliannejadi, G. Faggioli, N. Ferro, Vlachos, Michalis (Eds.), Working Notes of CLEF 2023–Conference and Labs of the Evaluation Forum, CLEF 2023, Thessaloniki, Greece, 2023.